Pedagogy for Higher Education Large Classes (PHELC26)
Online 12th June, 2026
DOI: 10.5281/zenodo.20606982

Scaling Up or Scaling Down? Rubrics, Peer Assessment and Feedback Literacy in a Large Postgraduate Cohort

Amara Atif https://orcid.org/0000-0003-0381-0343

Faculty of Engineering and Information Technology, University of Technology Sydney, Australia.

Abstract

This paper examines the use of rubrics to support peer assessment and feedback literacy in a large postgraduate cohort of over 210 students at an Australian university. Drawing on reflective practitioner inquiry, it explores how an analytic feedback rubric designed primarily as a feedback tool rather than a marking guide was deployed to scaffold peer assessment within a major group project. Three interrelated challenges are documented: inconsistent criterion application, surface-level compliance among a quarter of students, and affective barriers for culturally diverse learners. The paper reports on iterative adaptations, including co-developed rubric language and calibration activities, and reflects on their effectiveness. It argues that rubrics are necessary but insufficient for sustainable feedback at scale, and that their effectiveness depends on deliberate pedagogical scaffolding. The findings suggest that sustainable feedback at scale requires to be treated as dialogic tools rather than static artefacts.

Keywords: *Large classes; higher education; rubrics; peer assessment; feedback literacy.*

1. Introduction

Feedback is among the most influential factors in student learning (Hattie & Timperley, 2007), yet large classes in higher education create structural conditions that make meaningful feedback provision extremely difficult. The massification of higher education has intensified these pressures, with large classes posing well-documented challenges for engagement, quality feedback, and student learning outcomes (Hornsby & Osman, 2014; Jerez et al., 2021). In contexts such as this one, where enrolments exceed 100 students, the logistical demands of individualised, timely feedback become prohibitive, compelling educators to seek scalable alternatives. Rubrics, which describe performance criteria and standards, are among the most widely adopted of these alternatives (Panadero et al., 2023; Reddy & Andrade, 2010).

Rubrics take several forms. Holistic rubrics provide a single overall description of performance levels, enabling efficient judgements but limiting feedback specificity (Jonsson & Svingby, 2007). Analytic rubrics assess each criterion separately, generating more detailed, criterion-referenced feedback that helps students identify specific strengths and gaps (Andrade, 2000; Brookhart, 2013). Each type balances efficiency against specificity, a tension that becomes particularly acute at scale and shapes how rubrics are designed and deployed. A crucial distinction concerns whether rubrics function primarily as marking instruments or as feedback tools. In this paper, the rubric served both purposes but was designed with feedback as its primary intent, to shift students' orientation from receiving a score to identifying actionable next steps. This matters because students in large classes often engage with rubric-based feedback superficially, scanning for a grade rather than interrogating criteria (Boud & Molloy, 2013; Sharif & Atif, 2024; Taylor et al., 2024). Closing this gap requires building feedback literacy, defined by Carless and Boud (2018) as the capacity to seek out, make sense of, and act on feedback, through deliberate scaffolding including peer assessment as a vehicle for active criteria engagement (Topping, 2009).

This paper presents a reflective practitioner account of rubric design and deployment within a large postgraduate cohort of over 210 students. It examines how an analytic feedback rubric was used to scaffold peer assessment, the challenges that emerged, and the iterative refinements made in response, before drawing conclusions for large-class assessment design.

2. Teaching and Learning Context

The context for this paper is a postgraduate subject within an Information Technology and Business degree programme at a large metropolitan university in Australia. The subject introduces students to data analytics, dashboard design, and decision support, with a strong applied focus on real-world datasets and professional communication of insights.

Enrolments have grown to 210 students in the most recent offering. The cohort is characterised by considerable diversity: students include both domestic and international enrolments, represent multiple disciplines (IT, Business, and combined programmes), and demonstrate markedly varied prior experience with data and analytical tools. This diversity amplifies the challenge of differentiated feedback, as students arrive at the same assessment tasks with quite different backgrounds and support needs, a challenge that has been documented in similar large postgraduate cohorts at Australian universities (Lim et al., 2022).

Assessment comprises five components, the major element of which is a group project (40%) within which the peer assessment was embedded. Students were required to evaluate two peer group submissions against a structured analytic rubric and provide a minimum of 150 words of written commentary per dimension. The rubric covers two dimensions, Report Quality and Professional Presentation, each described across four performance levels (Developing, Functional, Proficient, and Exemplary) using behavioural descriptors drawn from professional standards in data communication and reporting.

Several interrelated challenges motivated the introduction of peer assessment and prompted subsequent reflection on rubric design. First, with over 210 students submitting group project artefacts, the volume of marking created a bottleneck that compressed turnaround times and reduced the quality of written feedback the teaching team could realistically provide. When feedback did arrive, it was often brief and generic, a consequence of scale rather than intent, but no less limiting in its impact on student learning.

Second, informal observation and student survey data indicated that a significant proportion of students were not reading or acting on the feedback they did receive. Several students in post-task consultations were unable to recall the criteria against which they had been assessed, suggesting that feedback was being treated as a terminal event rather than information to carry forward. This pattern is consistent with what Boud and Molloy (2013) describe as the single-loop feedback problem: students receive feedback but lack the orientation or scaffolding to use it productively.

Third, the diversity of student backgrounds created variability in how rubric language was interpreted. Terms such as “effective visualisation” or “coherent narrative” carried different meanings for students with different disciplinary preparation and cultural contexts, raising equity concerns where rubric descriptors rely on tacit professional knowledge that not all students can access equally. These challenges are not unique to this subject; they reflect structural conditions of large-class teaching in which feedback, even when provided at scale, may fail to produce the learning effects it is designed to generate (Nicol, 2010).

3. Literature Review

The literature on rubrics in higher education is extensive and broadly supportive of their use, particularly in large-class contexts where consistency and transparency are priorities (Ragupathi & Lee, 2020; Reddy & Andrade, 2010). Large classes present distinctive pedagogical challenges that the broader literature has begun to address systematically (Hornsby & Osman, 2014). One such challenge is maintaining consistent assessment standards across multiple markers, particularly acute when many tutors are marking student work against shared rubrics (Nicol, 2010). Rubrics have been shown to improve inter-rater reliability, reduce student uncertainty about expectations, and support self-regulation and self-efficacy (Jonsson & Svingby, 2007). A recent meta-analysis by Panadero et al. (2023) found a moderate positive effect on academic performance and a small but meaningful positive effect on self-regulated learning, findings that strengthen the case for rubric use in contexts where instructor feedback capacity is constrained. Taylor et al. (2024) further found that rubrics can increase students' confidence and reduce anxiety, particularly among students from non-traditional backgrounds for whom tacit academic conventions are less accessible. However, Panadero et al. (2023) also note that rubrics introduced as static documents shared once and left unmediated, produce weaker learning outcomes than those embedded in active pedagogical processes such as calibration, self-assessment, or peer review (Graham et al., 2022). Rubrics are therefore necessary but not sufficient; their impact depends substantially on how they are introduced, discussed, and enacted.

3.1. Rubrics and Feedback Literacy

Carless and Boud (2018) propose feedback literacy as a framework for understanding why feedback, including rubric-based feedback, so often fails to produce learning gains. They argue that effective feedback uptake requires students to develop four interconnected capacities: appreciating feedback, making judgements, managing affect, and taking action. In large classes, the structural conditions that support these capacities are typically difficult to create at scale. These include regular dialogue with instructors, low-stakes opportunities to practise applying criteria, and a psychologically safe environment for peer critique. Sadler's (1989) foundational work anticipates this argument: for feedback to close the gap between current and desired performance, students must hold an internal conception of the standard required, not merely receive a description of it. Rubrics can communicate the standard but cannot, by themselves, build that internalised understanding. Both perspectives point to the same conclusion: rubrics are most effective not as delivery mechanisms for standards, but as tools for building the capacity to understand and apply them.

3.2. Peer Assessment as a Vehicle for Rubric Engagement

Peer assessment has attracted growing interest as a strategy for large classes precisely because it distributes the feedback load while positioning students as active participants in the assessment process (Topping, 2009). When students apply a rubric to evaluate a peer's work, they engage with criteria actively rather than passively, a process that can deepen understanding of standards and improve the quality of their own subsequent work (Falchikov, 2005). Tai et al. (2018) argue that developing evaluative judgement, the capacity to assess the quality of work against shared standards, is central to students' long-term feedback literacy and is best cultivated through structured practice in making and justifying evaluative decisions. Gyamfi et al. (2022) demonstrated through a randomised controlled experiment that rubric use in peer assessment contexts significantly strengthens evaluative judgement over time. Liu and Carless (2006) identify key conditions for productive peer assessment: criteria must be comprehensible; students require preparation in making evaluative judgements; and the process must be integrated into subject design. Analytic rubrics are generally preferred for peer assessment in large classes because they provide more granular guidance, reduce cognitive load on assessors, and generate more specific and actionable feedback (Jonsson & Svingby, 2007; Nicol et al., 2014). This body of evidence suggests that peer assessment is most productive when rubrics are well designed, students are adequately prepared to apply them, and the process is embedded in subject design rather than treated as a standalone task.

4. Empirical Basis

This paper draws on reflective practitioner inquiry (Schon, 1983). Evidence informing the reflection includes an end-of-task student survey administered via the institutional learning management system (LMS; n=147, 70% response rate), LMS timed activity logs recording time spent on the peer review task, observations from drop-in student consultations following task submission, and an anonymous Mentimeter survey administered under institutional license across tutorial streams and aggregated for analysis. This survey followed the Week 7 calibration exercise in which students assessed a sample submission as a class activity. The use of LMS activity log data to identify engagement patterns is consistent with prior work in this context (Atif et al., 2020).

Survey data were analysed descriptively; open-ended responses were reviewed thematically, with recurring patterns identified through iterative reading. LMS activity log data were examined to identify students who completed the peer review in under five minutes, considered insufficient time for meaningful engagement with the task. Mentimeter responses were collected during the calibration sessions across tutorials and reviewed for emerging patterns, then compared against subsequent survey data. These sources were used

to triangulate observations and identify persistent challenges across offerings. No formal ethical approval was required as data were collected as part of routine subject evaluation processes. The findings are contextually situated and not generalisable, but are offered as a basis for practitioner reflection and dialogue consistent with the aims of this symposium.

5. Analysis and Reflection on Practice

The peer assessment process was introduced in Week 5, coinciding with the formal commencement of the major project, when students had sufficient foundational knowledge to engage meaningfully with the assessment criteria. Students were allocated two peer group submissions via the LMS and asked to complete the analytic rubric and provide written commentary per dimension. Peer grades contributed 20% of the task score, with instructor moderation applied where peer scores deviated substantially from the instructor's holistic assessment.

5.1. What Worked

Several aspects of the implementation yielded positive outcomes. Post-task survey responses (n=147, 70% response rate) indicated that 74% of students (n=109) agreed or strongly agreed that completing the peer review had helped them better understand the assessment criteria. Qualitative comments suggested that having to apply criteria to another's work had made the descriptors more concrete: one student noted they had not fully understood what "professional presentation" meant until they had to assess someone else against it. This is consistent with Topping's (2009) argument that peer assessment produces learning gains for reviewers as well as recipients, and with Gyamfi et al. 's (2022) finding that rubric use in peer assessment strengthens evaluative judgement over time. The written commentary produced was also noticeably more specific and criterion-referenced than in earlier peer review tasks using simpler rating scales, consistent with Jonsson and Svingby's (2007) finding that well-designed analytic rubrics improve peer feedback quality.

5.2. What Did Not Work

Significant challenges also emerged. First, despite the analytic structure, students applied criteria inconsistently: the same work received markedly different scores from different peer reviewers on the same dimension. This suggests that the descriptors, while detailed, remained open to varied interpretation, a problem that calibration activities might have addressed, but which class size made logistically difficult to implement at the outset.

Second, approximately 25% of students (n=37), identified through LMS activity log data, completed the rubric in under five minutes, insufficient time to read and evaluate a multi-page report carefully. These students tended to cluster scores in the mid-range and produced generic commentary that met minimum requirements without engaging

meaningfully with the criteria. This suggests that for a significant minority, the peer review process had become a compliance exercise rather than a genuine learning activity.

Third, students from culturally diverse backgrounds, particularly those for whom direct critical feedback is less normative, reported higher levels of discomfort with the peer reviewer role. This was evident both in survey responses and in drop-in consultations following task submission, where several students sought clarification on whether it was acceptable to give a peer a low score. This suggests that the rubric's technical scaffolding had not fully addressed the affective and relational dimensions of peer assessment that Carless and Boud (2018) identify as central to feedback literacy.

5.3. Iterative Adaptations

In response to these challenges, several adaptations were made. A calibration exercise was introduced in the Week 7 tutorial: students assessed a sample anonymised report as a class activity, with discussion facilitated around divergent scores. This timing was deliberate: students had sufficient project experience to engage meaningfully with the criteria but were early enough to adjust their approach. Mentimeter responses collected across tutorial streams (n=210) indicated that students felt more confident applying the criteria and understanding the standards expected following the calibration activity.

The rubric language was also revised iteratively, with technically ambiguous terms replaced by more explicit behavioural descriptions co-developed with students in a brief in-class consultation. Evaluative terms such as "clear" or "effective" were replaced by specific observable behaviours aligned to professional conventions, a change that reduced interpretive variability in subsequent offerings.

These adaptations reflect a broader principle: rubrics should be understood as dialogic tools rather than static artefacts. In large classes, the temptation is to deploy rubrics as finished products written once, embedded in the LMS, and left to do the work of feedback alone. The question for large-class educators should therefore not simply be "do we have a rubric?" but "what are we doing to ensure students can use it?". This aligns with evidence that effective learning design in large classes requires deliberate alignment between pedagogical structure and data-informed insights about student engagement (Lim et al., 2023). Sustainable feedback at scale requires legible rubrics, criteria calibrated through shared examples, and processes that position students as active participants.

6. Conclusion

This paper has argued that rubrics, while logistically essential in large-class contexts, are a necessary but insufficient mechanism for sustainable feedback at scale. The challenges documented here, inconsistent criterion application, surface-level compliance, and affective

barriers for culturally diverse students, reflect structural conditions common to mass higher education, where the scale that makes rubrics attractive also makes the relational work of building feedback literacy more difficult. Their value depends on deliberate scaffolding: calibration activities, concrete and co-developed language, and pedagogical structures that position students as active participants rather than passive recipients. A further challenge, not fully resolved here, concerns consistency among tutors who mark at scale: where multiple staff apply shared rubrics, the same calibration logic that benefits students must arguably extend to marking teams. Treating rubrics as living tools developed with students, rather than static artefacts deployed at them, remains both the key lesson from this study and a broader challenge for large-class assessment design.

References

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13-18. Retrieved from https://www.ascd.org/el/articles/using-rubrics-to-promote-thinking-and-learning

Atif, A., Richards, D., Liu, D., & Bilgin, A. A. (2020). Perceived benefits and barriers of a prototype early alert system to detect engagement and support 'at-risk' students: The teacher perspective. *Computers & Education, 156*, 103954. https://doi.org/10.1016/j.compedu.2020.103954

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698-712. https://doi.org/10.1080/02602938.2012.691462

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading.* Alexandria, VA: ASCD.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education.* London, England: Routledge. https://doi.org/10.4324/9780203220993

Gonzalez, J. (2014). Know your terms: Holistic, analytic, and single-point rubrics. *Cult of Pedagogy.* Retrieved from https://www.cultofpedagogy.com/holistic-analytic-single-point-rubrics/

Graham, A. I., Harner, C., & Marsham, S. (2022). Can assessment-specific marking criteria and electronic comment libraries increase student engagement with assessment and feedback? *Assessment & Evaluation in Higher Education, 47*(7), 1071-1086. https://doi.org/10.1080/02602938.2021.1986468

Gyamfi, G. B., Hanna, E., & Khosravi, H. (2022). The effects of rubrics on evaluative judgement: A randomised controlled experiment. *Assessment & Evaluation in Higher Education, 47*(1), 126-143. https://doi.org/10.1080/02602938.2021.1887081

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hornsby, D. J., & Osman, R. (2014). Massification in higher education: Large classes and student learning. *Higher Education, 67*(6), 711-719. https://doi.org/10.1007/s10734-014-9733-1

Jerez, O. M., Orsini, C., Ortiz, C., & Hasbun, B. A. (2021). Which conditions facilitate the effectiveness of large-group learning activities? A systematic review of research in higher education. *Learning: Research and Practice, 7*(2), 147-164. https://doi.org/10.1080/23735082.2020.1871062

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144. https://doi.org/10.1016/j.edurev.2007.05.002

Lim, L., Atif, A., & Farmer, I. (2022). 'Made good connections': Amplifying teacher presence and belonging at scale through learning design and personalised feedback. *ASCILITE Conference Proceedings,* 1-10. https://doi.org/10.14742/apubs.2022.55

Lim, L.-A., Atif, A., Heggart, K., & Sutton, N. (2023). In search of alignment between learning analytics and learning design: A multiple case study in a higher education institution. *Education Sciences, 13*(11), 1-21. https://doi.org/10.3390/educsci13111114

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279-290. https://doi.org/10.1080/13562510600680582

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education, 35*(5), 501-517. https://doi.org/10.1080/02602931003786559

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102-122. https://doi.org/10.1080/02602938.2013.795518

Panadero, E., Jonsson, A., Pinedo, L., & Fernandez-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review, 35*, 113. https://doi.org/10.1007/s10648-023-09823-4

Ragupathi, K., & Lee, A. (2020). Beyond fairness and consistency in grading: The role of rubrics in higher education. In C. S. Sanger & N. W. Gleason (Eds.), *Diversity and inclusion in global higher education* (pp. 73-95). Palgrave Macmillan. https://doi.org/10.1007/978-981-15-1628-3_3

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435-448. https://doi.org/10.1080/02602930902862859

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144. https://doi.org/10.1007/BF00117714

Schon, D. A. (1983). *The reflective practitioner: How professionals think in action.* Basic Books.

Sharif, H., & Atif, A. (2024). The evolving classroom: How learning analytics is shaping the future of education and feedback mechanisms. *Education Sciences, 14*(2), 1-15. https://doi.org/10.3390/educsci14020176

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467-481. https://doi.org/10.1007/s10734-017-0220-3

Taylor, B., Kisby, F., & Reedy, A. (2024). Rubrics in higher education: An exploration of undergraduate students' understanding and perspectives. *Assessment & Evaluation in Higher Education*, *49*(6), 799-809. https://doi.org/10.1080/02602938.2023.2299330

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20-27. https://doi.org/10.1080/00405840802577569